

Scientific report

on the project entitled:

Benchmark calculations on the components of QM/MM and fragment based quantum chemical method for describing biological systems

1. Motivation

This grant has been awarded since our paper (Kánnár, D., & Szalay, P. G. (2014). Benchmarking Coupled Cluster Methods on Valence Singlet Excited States. *Journal of Chemical Theory and Computation* **10**, 3757–3765. <http://doi.org/10.1021/ct500495n>) received particular attention from the community in the first two years after its publication. The present project has been designed to continue the benchmarking of quantum chemical methods describing excited states, including more properties, not just excitation energy. These properties include the so called Franck-Condon forces, the shape of the potential energy surface, dipole and transition moments. We also planned to investigate the effect of the environment on the excitation energies.

2. Results

The envisioned research has been divided into work packages (WPs). This allows us to summarise the accomplished research work according to the WPs.

Right at the beginning of the project there was a slight deviation from the original plans since I have received an invitation to participate in writing a Chemical Review paper on the methodology on calculation of excited states. My task was to prepare a chapter on Coupled Cluster methods, a topic closely related to the present project and which I am an expert of. (Lischka, H., Nachtigallova, D., Aquino, A. J. A., Szalay, P. G., Plasser, F., Machado, F. B. C., & Barbatti, M. (2018). Multireference Approaches for Excited States of Molecules. *Chemical Reviews* **118**, 7293–7361. <http://doi.org/10.1021/acs.chemrev.8b00244>). Note that Chemical Reviews is the highest impact journal in chemistry. The first year's report was accepted, acknowledging this detour.

WP1: This working package aimed at the characterisation of excited state potential energy surfaces by the Franck-Condon gradients and by a one-dimensional cut of the potential energy surface along the corresponding forces. Using the CCSD level as the standard approximation, we have investigated higher level methods including triple excitations which are more accurate, as well as lower level (so called second order) methods which are more economical. While we could show the importance of triple excitations on these properties, and identify an approximate non-iterative method (CCSD(T)(a)*) as the most cost effective but still accurate way of including triple excitations, we have concluded that the popular CC2 method fails sometimes badly, depending on the type of the excited state (Tajti, A., Stanton, J. F., Matthews, D. A., & Szalay, P. G. (2018). Accuracy of Coupled Cluster Excited State Potential Energy Surfaces. *Journal of Chemical Theory and Computation* **14**, 1–11. <http://doi.org/10.1021/acs.jctc.8b00681>). Although at this point we have completed the goals of WP1, the failure of CC2 encouraged us to look for more reliable alternatives. This is important from the point of view of large-scale applications, since higher order methods are too expensive to be applied for, e.g. biological systems. For example, our other research project (NKFI project K-124018) also relies on an economic electronic structure model. It turned out that the so called spin-component scaling technique can particularly cure the problem, resulting in CCSD quality results, and the corresponding SOS-CC2 and SOS-ADC(2) methods are even cheaper than the popular CC2. These results have been published recently in two papers (Tajti, A., & Szalay, P. G. (2019). Accuracy of Spin-Component-Scaled CC2 Excitation Energies and Potential Energy Surfaces. *Journal of Chemical Theory and Computation* **15**, 5523–5531. <http://doi.org/10.1021/acs.jctc.9b00676>, Tajti, A., Tulipán, L., Szalay, P. G. (2019). Accuracy of Spin-Component-Scaled ADC(2) Excitation Energies and Potential Energy Surfaces. *Journal of Chemical Theory and Computation*, [submitted for publication](#)).

WP2: This work package aimed at benchmarking quantities that are responsible for describing the interaction of distant chromophores. The most important quantity responsible for about 95% of the interaction is the dipole moment, which enters the interaction term via the transition dipole. Therefore, we have calculated the transition dipole at the Franck-Condon geometry and along the one-dimensional potential energy curves constructed in WP1 using CCSD, triples, as well as second order methods. Unfortunately, the results are moderately interesting: while the absolute value of the transition dipole depends on the method used, its relative change with nuclear coordinates runs rather parallel, independently of the theoretical model.

Therefore all methods seem to be appropriate describing the geometry dependence of this quantity. Thus, in our opinion, the results are not publishable in a separate article. Instead, it will be documented in a BSc thesis (Bónis Barcza), and the conclusion will be included in later publications describing the methodology of calculating chromophore-chromophore interactions (see again NKFI project K-124018).

WP3: In this WP benchmarking of methods was foreseen when also environmental effects are present. The effect of the solvent environment on the excitation energies was first investigated by solvating the formamide molecule. In the quantum chemical calculations the solvent molecules (in this case water) were represented by point charges replacing the atoms, taken from the TIP3P model (Jorgensen et al., *J Chem Phys*, **79**, 926 (1983)): -0.834 and 0.417 atomic units on O and H atoms, respectively. The configurations of the water molecules were generated by running a classical molecular dynamics simulation at 298K assuming NVT ensemble. The system consisted of the solute molecule represented by the OPLS force field (Jorgensen et al. *JACS* **110**, 1657 (1988)) and 247 water molecules in a cube-shaped periodic box with a dimension of 19.341 Å. The simulation was run for 10 ps, and snapshots were taken in every 100 fs resulting in 100 different structures. During the dynamics the structure of the formamide molecule was fixed in a geometry corresponding to the gas phase MP2 / cc-pVDZ energy minimum. The initial structure of the trajectory was a minimized energy configuration of the system corresponding to the above force field parameters. At each of the 100 different structures the excitation energies of the lowest-energy excited states were calculated with the EOM-CCSD method using the cc-pVDZ basis set (core electrons frozen), with the water molecules represented by point charges as described above. The resulting wavefunctions were assigned based on the overlaps with the excited state wavefunctions of the isolated molecule, and the results were statistically analysed. The following table summarises the main results:

State	Expectation value of the excitation energy / eV	Standard deviation of the excitation energy / eV
2A'	7.98	0.20
3A'	9.52	0.80
1A''	6.45	0.50
2A''	7.79	0.59

The magnitude of the standard deviation of excitation energies for the different snapshots is a surprising result. Since the spread is larger than the usual difference of the excitation energies obtained with different methods, further benchmarking by this technique does not seem to be reasonable; therefore we have terminated this WP.

WP4: In this work package we have constructed a benchmark set for dimers in order to test different quantum chemical methods for charge transfer (CT) states. The manuscript is in preparation, and will be submitted within a couple of weeks. This paper is a niche since no systematic study on the performance of electronic structure methods on CT states has been published yet. The results are also important, as we have found that the popular CC2 method fails again, underestimating the excitation energy of CT states by as much as 0.5 eV. Since for local states the average error is close to zero, in many cases even the order of the excited states comes out wrong. On the other hand, the above mentioned non-iterative triple method (CCSD(T)(a)*) is very accurate again, and also the standard method (CCSD) seems to quite reliable. To demonstrate the results, here we show one of the tables of the manuscript in preparation:

Table 4: Statistics on the error of excitation energies (in eV) for Charge Transfer states (CT character over 0.5) relative to CCSDT-3 reference values.

	mean	SD	min	max
CC2	-0.36	0.25	-0.76	-0.10
CCSD	0.26	0.10	0.10	0.44
CCSD(2)	0.29	0.16	0.05	0.53
CC3	-0.11	0.04	-0.18	-0.06
CCSD(T)(a)*	-0.02	0.03	-0.10	0.03
STEOM-CCSD	-0.06	0.13	-0.27	0.23
Bt-PNO STEOM-CCSD	-0.07	0.13	-0.27	0.23

It remains to be seen whether spin-component scaling can cure the error of CC2. In the meantime the results of this WP have been presented as a poster (Kozma, B., Tajti, A., Szalay, P.G.: Benchmark calculations with Coupled Cluster methods on Charge Transfer states) at the *CESTC 2019 conference* and was in part compiled in a BSc thesis (Balázs Kozma).

3. Summary

We believe that during the short term of this project we have made substantial progress on the field of benchmarking Coupled-Cluster type methods for excited states. First of all, the scope of such tests has been extended beyond the vertical excitation energies. We have produced a large amount of reference data to test methods, which allow one to perform tests not only on vertical excitation energies, but on other properties as well. In fact, the results show that this is rather important, since good vertical excitation energies do not guarantee that potential energy surfaces or other properties will be reliably modelled by the given quantum chemical method. Fortunately, we could identify a cheap second order method (SOS-ADC(2)) that seems to work very well in many applications. This finding not only encourages the use of this simplified method, but also opens up the possibility of reliable modelling of large, even biological system.

Four papers have been published in Q1 journals already, and a fifth one is in preparation. We believe that the community will appreciate these results, as did our 2014 paper mentioned above. Non-published additional results are also available which, as planned, will be used in our forthcoming research, as for example in our other NKFI project K-124018, and can be published as essential ingredients of the forthcoming manuscripts.